

Bayesowski Model Normalnej Regresji Liniowej – ćwiczenie

- 1) Zapoznaj się z danymi zestawionymi w pliku *Wyplaty_gotowki.xlsx*
- 2) Niech narzędziem modelowania zależności pomiędzy wartością wypłat gotówki w bankomatach w Polsce a liczbą bankomatów raz wydanych kart płatniczych będzie model regresji postaci:

$$W_t = \beta_0 + \beta_1 t + \sum_{i=2}^4 \beta_i Q_{ti} + \beta_5 B_t + \beta_6 K_t + \varepsilon_t, \quad \varepsilon_t \sim iN(0, \tau^{-1}),$$

gdzie Q_{ti} ($i = 2, 3, 4$) reprezentują zero-jedynkowe zmienne sezonowe odpowiadające kwartałom II ($i = 2$), III ($i = 3$) i IV ($i = 4$). Zbuduj stosowne macierze y oraz X , odpowiadające macierzowemu zapisowi tego modelu: $y = X\beta + \varepsilon$

Dla powyższego modelu przeprowadź estymację bayesowską przyjmując gamma-normalny rozkład *a priori* z zależnością pomiędzy β i τ , przy założeniu, że:

- $E(\tau) = 1, Var(\tau) = 0,5$
- $E(\beta) = 0_{(k \times 1)}$ i $V(\beta) = 100I_k$ (zatem $Var(\beta_i) = 100, i = 1, 2, \dots, k$, oraz $Cov(\beta_i, \beta_j) = 0, i \neq j$)

Dla poszczególnych parametrów (tak strukturalnych, jak i precyzji składników losowych) wyznacz następujące **charakterystyki** (te w brzegowych rozkładach *a posteriori* zestaw z tymi *a priori*):

- wartości oczekiwane, mediany, modalne
- odchylenie standardowe
- przedziały najwyższej gęstości *a posteriori* (dla poziomu prawdopodobieństwa 0,95)

Dodatkowo, dla parametru odzwierciedlającego związek pomiędzy liczbą bankomatów w wartości wypłat gotówkowych (i tylko dla niego) **skonstruuć wykres** przedstawiający gęstości rozkładów *a posteriori* i *a priori* (na tym samym wykresie).

Uwaga 2: W celu ustalenia zakresu wartości na osi OX tego wykresu, możesz przyjąć regułę, że zakres ten stanowi przedział $(E(\beta_5|y) - 4D(\beta_5|y), E(\beta_5|y) + 4D(\beta_5|y))$, z końcami zaokrąglonymi do np. 2-3 miejsc po przecinku. Przedział ten można następnie podzielić np. na 100 lub więcej równych odcinków, w ten sposób wyznaczając w arkuszu „siatkę punktów” (w postaci kolumny) wartości parametru, w których zostaną obliczone wartości funkcji gęstości rozkładów *a posteriori* i *a priori* (w centrum wykresu będzie się znajdował rozkład *a posteriori*, zaś to, na ile rozkład *a priori* będzie widoczny na tym wykresie, zależeć będzie od tego, jak daleko jest on ulokowany w stosunku do rozkładu *a posteriori*). W każdym punkcie tej „siatki” należy obliczyć wartości funkcji gęstości stosownych rozkładów.

Uwaga 3: O ile wartości funkcji gęstości rozkładu gamma można obliczyć w MS Excel za pomocą dostępnej tam funkcji, o tyle nie ma w arkuszu funkcji obliczającej wartość gęstości 1-wymiarowego rozkładu t-Studenta $St(m, p, v)$, gdzie m oznacza parametr niecentralności (modalną), p – precyzję, zaś v – liczbę stopni swobody ($v > 0$). Dlatego wartości funkcji gęstości tego rozkładu należy obliczyć wprowadzając „ręcznie” do komórki wzór:

$$f_{St}(x|m, p, v) = \frac{\Gamma(\frac{v+1}{2})\sqrt{p}}{\Gamma(\frac{v}{2})\sqrt{v\pi}} \left(1 + \frac{p(x-m)^2}{v}\right)^{-\frac{v+1}{2}}, \quad x \in \mathbb{R}$$

- 3) Zweryfikuj statystyczną istotność poszczególnych parametrów strukturalnych.